Guidelines for Housing IT Infrastructure

Information Item

Introduction

The Technology Services Board requested that the Department of Technology Services (DTS) propose guidelines for data centers and server rooms that house information technology (IT) infrastructure. This request was initiated due to various studies currently underway that address existing IT data centers and server rooms as part of IT infrastructure consolidation efforts. Additionally, the server consolidation report presented to the Board by Intel Corporation recommended various steps to consolidate certain IT infrastructures.

Background

Data Center design must consider a variety of factors including condition, capacity, and capability. More recently, this list has been expanded to include Green (the goal of achieving sustainable environmental objectives such as carbon footprint reductions). Recognizing these factors during the initial design of a facility can improve service levels, security, reliability, and reduce overall costs as well as achieving environmental goals.

Data Centers and Server Rooms Guidelines and Standards

The Uptime Institute provides detailed specifications for four (4) tiers of IT infrastructures as follows:

Tier I and II facilities have limited redundancy and are typically chosen to address short-term housing needs. Applications housed in Tier I/II facilities are primarily driven by cost and time to market objectives. Tier III/IV facilities have built in redundancies and fault tolerances. These are considered strategic solutions chosen for rigorous uptime and long-term needs. It has been found that it is cost prohibitive to upgrade a facility by more than one tier. Fiscal prudence dictates building the desired tier level initially, thus avoiding expensive upgrades and related service outages.

Conforming to these tier standards will normally result in facility uptime of between 99.67% and 99.99% (Ignoring other component layers that contribute to overall availability).

Additionally, the condition, capacity, and capability can be rated as good, average, or poor based on such factors as redundancy, floor space conditions, and fire protection.

Green standards also exist. The U.S. Green Building Council's Leadership in energy and Environmental design (LEED) has developed a nationally accepted benchmark for buildings ranging from silver to platinum. The State's current standard for buildings is silver.

Data Centers and Server Rooms Guidelines

January 2008

Table of Contents

1. Introduction

Clark Kelso, former California State Chief Information Officer, requested that the Department of Technology Services (DTS) propose standards and guidelines for data centers and server rooms that house information technology (IT) infrastructure, and present them to the Technology Services Board (TSB). This request was initiated, in part, due to the California Department of Transportation (Caltrans) Data Center Assessment, which addressed the condition of their existing IT data centers and server rooms as part of their IT infrastructure consolidation effort. This guideline is the product resulting from that request.

> Note*:* The request is relative to both data centers and server rooms. Combined, data centers and server rooms are often referred to as data center equivalents (DCEs).

Agencies who conduct in-house assessments of their aging data centers and server rooms will also need to conduct an assessment of their IT systems, their data and their availability requirements.

At the lower end of the spectrum is a “sandbox” or development server for a small application. The need to protect data is low; availability is not 24 x 7; uptime is not critical; and recovery is not urgent. For this example, a Tier III Data Center is not required and may not be cost-effective. A secure, on-site server room may be sufficient to meet the server environment needs.

At the higher end of the spectrum is an email enterprise server for the agency. The need to protect data is critical; availability is 24 x 7; uptime is critical; and recovery is critical. For this example, a Tier III Data Center is highly recommended.

A Tier III Data Center provides physical security, network and application security, fire protection, fire suppression, emergency services, redundancy, 24 x 7 availability and recovery.

> *Note:* If there is no critical data processing (e.g., development and testing environments, or temporary “sandbox” environments), the design requirements are different from a DCE that hosts a mission critical solution (e.g., email).

2. Background

In early 2004, the Caltrans IT division began a Consolidation Project that encompassed the unification of IT resources within Caltrans. The objective of the Consolidation Project was to improve effectiveness resulting in standardized environments, equipment, architecture, and staffing skills directed toward achieving savings in staffing resources, software, hardware and facilities. In early 2007, one of the first of the affiliated projects related to consolidation was an assessment of the Caltrans DCEs.

Throughout California state government organizations, many aging data centers and server rooms do not meet industry standards for a modern data center facility design. The most common DCE characteristics are often grouped as ***Condition***, ***Capacity*** and ***Capabilities***.

> *Note:* Recently, this has been expanded to include ***Green***, the goal of achieving sustainable environmental objectives such as carbon footprint reductions.

3. Drivers

The drivers to create, adopt, implement and sustain standards and guidelines for DCEs can be grouped into 4 major categories:

Condition – The overall condition is based upon the environment, including elements such as heating, ventilating and air conditioning (HVAC), power, equipment cooling, fire suppression, premise cabling, physical security and redundant systems.

Capacity – The capacity is also based upon the environment, including the following considerations:

- Is the existing mechanical infrastructure adequate to support the current environment?
- Is it adequate to support the future environment (based on business needs)?
- Is there adequate floor space for the equipment hosted?
- Is there adequate fire protection?
- Is there adequate telecommunications?

Capability – Capability is related to the types of solutions supported (i.e., hosted). For non-mission critical solutions, the capability requirements are much lower than the capability requirements for mission critical solutions. The following areas are considered part of capability:

- Facility Connectivity
- Technical Infrastructure
- Emergency Services
- Backup and Restore
- Operational Recovery and Disaster Recovery

Green – Green, as it applies to buildings and especially data center facilities, is gaining momentum in both private and public sectors. It is both politically correct and financially good business to incorporate Green in the design, operation and maintenance of data centers and server rooms. While there have always been financial drivers for IT, acknowledging specific Green drivers have not been in the forefront. However, public pressure is increasing for going Green in IT. The following areas are considerations:

- Total energy consumption
- Electricity consumption
- CO_2 emissions
- Raw materials used
- Potable water consumed
- Waste output

A carbon footprint is a measure of the impact that activities have on the environment in terms of the amount of green house gases produced, measured in units of carbon dioxide". It is meant to be useful for organizations to conceptualize their impact in contributing to global warming.

4. Benefits

Creating, adopting and implementing DCE standards and guidelines will result in:

Improved service levels and accountability – Service levels benefit from good physical conditions. Service availability increases and capacity planning (e.g., equipment refresh and program growth) is more predictable. Facility managers, equipment managers and service managers can accept accountability with confidence in knowing that standards and guidelines have been implemented.

Increased security – Physical security is easier to implement and sustain than other elements of security. Increases in physical security result in increases in the overall security of the solutions hosted.

Increased reliability – As service availability and security increase, solutions become more reliable. Together, availability, security and reliability result in increased user and public confidence for mission critical solutions.

Reduced data center and server room costs – .Cost reductions will occur in many areas. For example, a proper raised floor requires less cooling, and often results in fewer equipment failures. Appropriate redundancies allow for operations and maintenance to occur concurrently, further reducing overall costs.

Leveraged resources and equipment through the use of common infrastructure – Use of common infrastructure (e.g., enterprise backup power, enterprise storage, emergency uninterrupted power supply, and shared servers) allows for less staff training/expertise. One-of-a-kind components are more difficult to sustain.

Increased financial savings through economies of scale – A properly managed DCE in good condition results in consolidation of equipment, resources and staff; and creates financial savings.

Achieved environmental goals – LEED® certification, which includes a rigorous third-party commissioning process, offers compelling proof to clients, peers and the public at large that environmental goals have been achieved and that the building is performing as designed.

5. Standards for Data Centers and Server Rooms

5.1. Availability Tiers of IT Infrastructure Facilities

The Uptime Institute™[1] provides detailed specifications for 4 tiers of IT infrastructure and loose decision criteria for selecting which tier best meets individual needs. Each tier consists of power, cooling and other critical physical environmental sub-systems that work together in a tightly integrated uptime system (these are itemized later in this document).

Conforming to these tier standards will typically result in a "facility" uptime accomplishment between 99.67% and 99.99% percent, ignoring the other component layers that contribute to overall availability attainment. The Uptime Institute™ has tracked failure statistics for 12+ years and has determined that it typically takes at least five and often seven facility interacting problems before a facility downtime failure occurs.

The higher the transition through the tier structure, the greater ability to conduct repairs and maintenance without impacting the services housed within the facility.

Note*:* The tier rating system for an entire site is limited to the rating of the weakest subsystem that will impact site operation.

TIER I - Basic Site Infrastructure

A Tier I facility has non-redundant capacity components (i.e. power source) and single non-redundant path distribution serving the site's computer equipment. Basically, any capacity component or distribution path failure will negatively impact the computer equipment, hence the service. In addition, any planned facility work will require most or all computer equipment to be powered off. Annual site maintenance will require the entire facility to be powered down. Average impact of maintenance and unplanned site failures is approximately 28.8 hours a year, or 99.67% uptime.

TIER II - Redundant Capacity Components Site Infrastructure

A TIER II facility has redundant capacity components (UPS and engine generators) and single non-redundant path distribution serving the site's computer equipment. Basically, any capacity component failure still has a chance of causing a site failure, while any distribution path failure will cause the computer equipment to shut down. Annual site maintenance is

[1] http://uptimeinstitute.org/component/option,com_frontpage/Itemid,1/

required that will require the entire facility to be powered down. Average impact of maintenance and unplanned site failures is approximately 22 hours a year, or 99.75% uptime.

TIER III - Concurrently Maintainable Site Infrastructure

A TIER III facility has redundant capacity components and multiple distribution paths serving the site's computer equipment. Generally, only one distribution path services the computer equipment at any given time. All computer equipment must have dual power inputs. Operational errors or spontaneous failures of site infrastructure may still cause a facility failure. Average impact of maintenance and unplanned site failures is approximately 1.6 hours a year, or 99.98% uptime.

TIER IV - Fault Tolerant Site Infrastructure

A TIER IV facility has redundant capacity systems and multiple distribution paths simultaneously serving the site's equipment. All computer equipment is dual powered and installed properly to be compatible with the topology of the site's facility architecture. A single worst-case failure of any capacity system, capacity component or distribution element will not impact the computer equipment's operation. Any component or element can be removed from service for maintenance, repair or replacement without impact the computer equipment's operation. Average impact of maintenance and unplanned site failures is less than 1 hour a year, or 99.99% uptime.

Note*:* Any incident involving fire alarms, fire suppression or initiating an emergency power off (EPO) will cause a data center disruption.

Note*:* Formal site certification is obtained through the Uptime Institute™

Note*:* The Uptime Institute™ does not rate using fractional tier ratings. The DCE tier rating is the lowest of the component ratings.

Note*:* The classification ratings assigned to the data center sites are based on opinions from experts within the industry that perform assessments.

5.2. Condition

Condition Rating	**Good:** Meets many best practices and has no critical risks.

Examples of components contributing to a good condition rating include:

- Some redundancy
- Floor space is in excellent condition
- Telecommunications access paths have multiple access points and multiple providers
- Fire protection is sufficient.

Condition Rating	**Average:** Meets some best practices but may have some areas of concern. No critical risks.

Examples of components contributing to an average condition rating include:

- Limited redundancy
- Floor space conditions are average
- Telecommunications access paths have single access points but multiple providers
- Fire protection is available but needs improvement.

Condition Rating	**Poor:** Facility contains multiple points of failure and has several risk areas.

Examples of components contributing to a poor condition rating include:

- Little or no redundancy
- Floor space conditions are poor
- Telecommunications access paths have single access points and/or single providers
- Fire protection is out-dated and/or insufficient for the facility

5.3. Capacity and Capability

Standards for capacity and capability are often grouped in 4 areas: electrical, mechanical, telecommunications and other.

Electrical – Standards for electrical capacity and capability are measureable in watts per square foot and UPS capacity in kilowatts of UPS capacity.

Mechanical - Standards for mechanical capacity and capability are measurable for various areas:

- Mechanical Infrastructure
- HVAC
- Equipment Cooling Capacity (measureable in tons per kilowatt)
- Equipment (configuration, redundancies, backup)
- Mechanical and Electrical Equipment Backup (generators)
- Fire Suppression (water or gaseous)
- Airflow (measurements)
- Heat Index (measurements)
- Workspace (measurements)

Telecommunications – Standards for telecommunications capacity and capability are measurable for various areas:

- Access Paths (multiple access points and multiple providers)
- Cable Plant Infrastructure and Pathways
- Backbone Cabling
- Points of Entry
- Types of cabling (copper or fiber)

Other – Standards for capacity and capability are also measurable for various areas not included above:

- Location
- Fire Protection
- Emergency Services
- Physical Security
- Data Center Administration
- Concurrent Operations and Maintenance
- Adequate Floor Space
- Raised Floor Height
- Single Points of Failure
- Ability to Restore Operations

Capacity and Capability Rating	**Good:** Meets many best practices and has no critical risks.

Examples of components contributing to a good capacity rating include:

- Able to support equipment for the existing loads
- Floor space has capacity for growth
- Telecommunications access paths have multiple access points and multiple providers
- Fire protection is sufficient

Examples of components contributing to a good capability rating include:

- Redundancy
- Sufficient capability to support equipment for the existing loads
- Able to perform concurrent operations and maintenance
- Floor space conditions are good
- Telecommunications access paths have multiple access points and multiple providers

Capacity and Capability Rating	**Average:** Meets some best practices but may have some concern areas. No critical risks.

Examples of components contributing to an average capacity rating include:

- Sufficient capacity to support equipment for the existing loads
- Floor space has some room for growth
- Telecommunications access paths have single access points but multiple providers
- Fire protection is available but needs improvement

Examples of components contributing to an average capability rating include:

- Limited redundancy
- Sufficient capability to support equipment for the existing loads
- Not able to perform concurrent operations and maintenance
- Floor space conditions are average and/or some capability for growth
- Telecommunications access paths have single access points but multiple providers
- Fire protection is available but needs improvement.

Capacity and Capability Rating	**Poor:** Facility contains multiple points of failure and has several risk areas.

Examples of components contributing to a poor capacity rating include:

- Limited capacity to support equipment
- Limited or outdated support equipment
- Floor space has no room for growth
- Telecommunications access paths have single access points
- Fire protection is insufficient for the facility

Examples of components contributing to a poor capability rating include:

- Little or no redundancy
- Not able to perform concurrent operations and maintenance
- Limited or outdated support equipment
- Floor space conditions are poor
- Telecommunications access paths have single providers
- Fire protection is out-dated for the facility

5.4. Green

The U.S. Green Building Council[2] (USGBC) has developed the Leadership in Energy and Environmental Design[3] (LEED®). LEED® is the nationally accepted benchmark for the design, construction and operation of high performance buildings. LEED® gives building owners and operators the tools they need to have an immediate and measurable impact on their buildings' performance. LEED® promotes a whole-building approach to sustainability by recognizing performance in five key areas of human and environmental health: sustainable site development, water savings, energy efficiency, materials selection and indoor environmental quality.

Architects, real estate professionals, facility managers, engineers, interior designers, landscape architects, construction managers, lenders and government officials all use LEED® to help transform the built environment to sustainability. State and local governments across the country are adopting LEED® for public-owned and public-funded buildings; there are LEED® initiatives in federal agencies, including the Departments of Defense,

[2] http://www.usgbc.org/
[3] http://www.usgbc.org/DisplayPage.aspx?CategoryID=19

Agriculture, Energy, and State; and LEED® projects are in progress in 41 different countries, including Canada, Brazil, Mexico and India.

LEED® for New Construction addresses six important environmental areas:

- Sustainable Sites
- Water Efficiency
- Energy & Atmosphere
- Materials & Resources
- Indoor Environmental Quality
- Innovation in Design

LEED® for New Construction Ratings[4]	
Platinum	**52-69 points**
Gold	**39-51 points**
Silver	**33-38 points**
Certified	**26-32 points**

[4] http://www.usgbc.org/ShowFile.aspx?DocumentID=1095

6. Guidelines for Data Centers and Server Rooms

6.1. Availability Classification

Tier I and II solutions are typically chosen to address short-term housing needs and driven primarily by cost and time to market objectives. These solutions are considered tactical and are not designed for high uptime or availability business needs.

Tier III and IV are considered strategic solutions and are typically chosen due to the rigorous uptime and long term nature of the business investment. These tiers are designed to have an effective life beyond the immediate computer equipment requirements. These tiers enable the business owner to make strategic business decisions concerning growth and technology, unconstrained by current site topology.

If the business has long term (strategic) plans for the facility, it has been found that it is difficult (cost prohibitive) to upgrade an existing facility by more than 1 tier. This means that upgrading from a Tier I to a Tier II is possible, but subsequently trying to upgrade that site to a Tier III is not practical. Therefore, if the business's long-term vision will ultimately require a Tier III facility, it may be fiscally prudent to build that from the onset rather than experience expensive upgrade costs and service outages due to construction/renovation.

Besides the above characteristics, the Uptime Institute™ provides examples of what type of business need is appropriately addressed by each tier.

Tier I is appropriate for business needs such as:

- Information Technology being deployed that primarily supports internal business processes.
- Use of "web-presence" solutions are passive, information only types.
- Companies without a quality of service commitment for their IT services.
- Only accessed during normal business hours.

Tier II is appropriate for business needs such as:

- Internet applications without serious financial penalties for quality of service commitments.
- Only operating during normal business hours and can be shut down during off hours.
- Not used to house on-line or real-time service delivery solutions.

Tier III is appropriate for business needs such as:

- Housing solutions that support internal and external customers with 7x24 needs, but can be shut down when limited service is acceptable.
- Housing solutions that support multiple time zones (www).
- Housing solutions supporting internal business processes that can tolerate system shutdowns.

Tier IV is appropriate for business needs such as:

- Housing solutions that support internal and external customers with 7x24 needs, but can not afford to be shut down.
- Housing solutions that are e-commerce, market transactions or financial settlement support.
- Housing services that span multiple time zones (www) and are mission critical.

6.2. Condition

The data center and server room condition rating takes into account several factors of the environment including mechanical capacity and redundancy, floor space capacity, power, points of failure, telecommunications, and fire protection.

- What is the condition of the existing mechanical infrastructure?
- What is the condition of the existing power and cooling systems?
- What is the condition of the premise cabling?
- Are there improvements that should be made to the existing infrastructure support equipment that would increase its overall reliability?

It is important to conduct an assessment on a regular basis. Areas of an assessment often focus on:

- Size of data center - number, type, location;
- Environmental - HVAC, power, equipment cooling;
- Physical layout - raised floor, overhead cabling, accessibility, height, fire detection; and
- Equipment - type, physical configuration, redundancies, and backup.

In addition, other areas related to the physical infrastructure include:

- Support Function;

- Facility Connectivity;
- Security;
- Technical Infrastructure;
- Data Center Administration; and
- Emergency Services.

6.3. Capacity and Capability

Guidelines for capacity and capability are often grouped in 4 areas: electrical, mechanical, telecommunications and other.

Electrical **–** Guidelines for electrical capacity and capability include adding centralized UPS units and maximizing modular UPS units. Implement redundancy where possible.

Mechanical **–** Guidelines for mechanical capacity and capability include: implement redundancy where possible for power, backup power, emergency power, HVAC, equipment, equipment cooling and fire suppression.

Telecommunications **–** Guidelines for telecommunications capacity and capability include adding Category 6a cabling and high density infrastructure components. Multiple access points and multiple providers should be used. Implement redundancy where possible.

Other **–** Other guidelines for capacity and capability include locating facilities in diverse geographic areas to mitigate the impact from natural disasters. Implement redundancy where possible. Consider and evaluate the overall location for fire protection, emergency services and public security.

Establish enterprise standards for data center operations and support. Address business continuity, service continuity, operational recovery and disaster recovery through formal efforts.

6.4. Green

Guidelines for Green are often grouped in 6 areas:

- Energy consumption
- Electricity consumed
- CO2 emissions and other Green House Gas (GHG) emissions

- Raw materials used
- Potable water
- Waste output.

Whether building new or remodeling, there are various ways to reduce GHG

- Monitor the current carbon footprint
- Identify potential improvements for energy consumption and associated CO2 emissions.
- Investing in projects that reduce CO2 emissions (e.g., solar energy, wind energy, biofuels, recycling, Green purchasing, reforestation, tree planting). This includes purchasing from vendors who “go Green”.

Appendix A: Uptime Calculations

Uptime can be measured in many different ways:
- Total infrastructure in a Data Center and Server Room.
- Individual components (e.g. equipment power and equipment cooling).
- A Service (e.g., email).
- A Service Components (e.g., servers and networks).

Uptime is often misunderstood (and therefore misused) and incorrectly communicated:
- It generally excludes normal and regularly scheduled maintenance windows.
- It generally includes emergency maintenance.
- It generally can be described for each hardware device.
- It is often misunderstood when two or more uptime calculations are involved.

Common Tier Availability:
- Tier I: 99.67% availability
- Tier II: 99.75% availability
- Tier III: 99.98% availability
- Tier IV: 99.99% availability

Uptime Calculations:
- A **99.67%** uptime has outages of **28.8 hours** in a **1**-year operating period.
- A **99.75%** uptime has outages of **22.0 hours** in a **1**-year operating period.
- A **99.98%** uptime has outages of **1.6 hours** in a **1**-year operating period.
- A **99.99%** uptime has 1 outage of **4.0 hours** in a **5**-year operating period.
- A **99.999%** uptime (commonly referred to as "Five Nines" or "24 by forever") has 1 outage of **4.0 hours** in a **45.6**-year operating period (45.6 years x 365 days x .99999% = 4.0 hours).

Example 1:

If one device has a 99.5% uptime, and a second device has a 99.5% uptime, it is often misunderstood to be that the two devices combined have a 99.5% uptime. In fact, the calculation is 99.5% x 99.5%, which equals 99%.

> Note*:* According to the Uptime Institute™, "…***70% or more of all site failures involve people. Of these failures, 2/3 are management error and 1/3 are caused by people..."***[5]

[5] http://uptimeinstitute.org/wp_pdf/TUI705DTierClassification_WP.pdf

Appendix B: Notes on Referenced Documents

Reference 1: Uptime Institute™:
"Tier Classifications Define Site Infrastructure Performance "[6]

Note*:* The Electrical System Topology Diagrams for Tiers I, II, III and IV appear on pages 14-17.

Reference 2: Uptime Institute™:
"Four Metrics Define Data Center 'Greenness'"[7]

Note*:* The four metrics are:

- IT strategy (business requirements, systems design and architecture, platform, data topology and network design)
- IT hardware asset utilization
- IT energy and power hardware efficiency
- Site physical infrastructure overhead

Reference 3: Uptime Institute™:
"A Simple Model for Determining True Total Cost of Ownership for Data Centers"[8]

Reference 4: Uptime Institute™:
"The Invisible Crisis in the Data Center: The Economic Meltdown of Moore's Law"[9]

Reference 5: Uptime Institute™:
"Data Center Energy Efficiency and Productivity"[10]

Reference 6: LEED®
"LEED® for New Construction & Major Renovations"

Reference 7: LEED®
"LEED® for Existing Buildings: Operations and Maintenance"

[6] http://uptimeinstitute.org/wp_pdf/TUI705DTierClassification_WP.pdf
[7] http://uptimeinstitute.org/wp_pdf/(TUI3009F)FourMetricsDefineDataCenter.pdf
[8] http://www.uptimeinstitute.org/wp_pdf/(TUI3011C)SimpleModelDetermingTrueTCO.pdf
[9] http://uptimeinstitute.org/wp_pdf/(TUI3008)Moore'sLawWP_080107.pdf
[10] http://uptimeinstitute.org/wp_pdf/(TUI3004D)DataCenterEnergyEfficiency.pdf